Warszawa, 21 stycznia 2022 r.

dr hab. inż. Robert Nowak, prof. uczelni
Instytut Informatyki
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska
ul. Nowowiejska 15/19
00-665 Warszawa

Recenzja rozprawy doktorskiej mgr Julii Herman-Iżyckiej zatytułowanej „Analiza różnicowej ekspresji genów w populacjach bakterii"

Recenzja powstała na prośbę Rady Naukowej Dyscyplin Matematyka i Informatyka Uniwersytetu Warszawskiego na podstawie: (1) poprawionej rozprawy doktorskiej z grudnia 2021 r, (2) autoreferatu, (3) dorobku naukowego Kandydatki uwzględnionego w bazach Scoupus i Google Scholar, (4) kodów źródłowych prezentowanych rozwiązań dostępnych w repozytorium GitHub, (5) poprzedniej mojej recenzji z 16 sierpnia 2021 r., (6)odpowiedzi na recenzje mgr Julii Herman-Iżyckiej, (7) pisma z wyjaśnieniami promotora - dr hab. Bartosza Wilczyńskiego.

1 Tematyka badań

Rozprawa dotyczy metod analizy danych metatranskryptomicznych (sekwencji RNA) uzyskiwanych przez sekwencjonowanie nowej generacji metodą Illumina. Tytuł rozprawy odpowiada jej treści. Mgr Julia Herman-Iżycka zaproponowała nowe metody analizy tego typu danych pozwalających badać aktywność mikroorganizmów (głównie bakterii) zamieszkujących daną niszę środowiskową. Metody te wykorzystano dla nowych, rzeczywistych danych. Rozprawa definiuje problem asemblacji różnicowych kontigów, pokazuje nowe algorytmy oraz ich realizacje w języku Python.

Cel badawczy jest postawiony właściwie, jest on interesujący i istotny. Przedstawione rozwiązania są poprawne i potwierdzone eksperymentem.

2 Główne wyniki rozprawy

W rozprawie w sposób właściwy przeprowadzono analizę źródeł z literatury światowej, co świadczy o dostatecznej wiedzy Autorki. Wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący, pozwoliły one postawić opisane w pracy problemy badawcze.

Mgr Julia Herman-Iżycka użyła właściwej metody do rozwiązania postawionych problemów.

W rozprawie pokazano rozwiązanie dwóch spójnych tematycznie problemów. Pierwszym z nich, opisanym w rozdziale 3, jest uzyskanie ilościowej oceny zmian w mikrobiomie jelitowym myszy na różnych dietach. Istniejące metody pozwalały wykorzystać

tylko niewielką część sekwencji. Zespół, którego członkiem jest Autorka rozprawy zaproponował badanie różnic ekspresji genów, wykorzystanie bazy genomów referencyjnych i nowe algorytmy wykorzystujące m.in. asemblację. Pozwoliło to w ciekawy sposób opisać w jaki sposób mikroorganizmy żyjące w przewodzie pokarmowym myszy oraz organizm gospodarza oddziałują na siebie.

Drugim problemem, opisanym w rozdziale 4, jest uzyskanie metody, która wykorzystując odczyty z sekwenatorów jednocześnie tworzy sekwencję wynikową (problem asemblacji) i identyfikuje sekwencje, które różnicują dwie grupy próbek (problem identyfikacji). Zagadnienie takie występuje m.in. przy analizach odczytów z mikrobiomów. Kandydatka przeprowadziła rozważania teoretyczne dla struktur danych wykorzystywanych w asemblacji: pod-grafów grafów de Bruijna oraz dla grafów nałożeń, zaproponowała nowy algorytm i go zrealizowała jako moduł programu String Graph Assembler (SGA). Porównała wyniki z innymi rozwiązaniami i wyprowadziła właściwe wnioski.

Uzyskane wyniki są pokazane poprawnie. Wywód jest zwięzły i jasny, drobne nieścisłości przedstawiłem w poprzedniej recenzji. Doktorantka poprawnie odniosła się do poprzednich uwag i nieścisłości uzupełniając rozprawę. Nie stwierdzam znaczących błędów redakcyjnych w obecnej wersji.

3 Uwagi krytyczne

Uwagi krytyczne zostały wymienione w poprzedniej recenzji, z dn. 16 sierpnia 2021 r., większość z nich doktorantka uwzględniła w aktulanej wersji rozprawy. Chciałbym poruszyć jeszcze dwie kwestie.

1. Moim zdaniem wadą rozprawy jest brak publikacji rozwiązań przedstawionych w rozprawie w czasopiśmie lub w materiałach konferencyjnych. Autorka w autoreferacie wspomina, że publikacje takie są w przygotowaniu.

2. Przegląd metod asemblacji jest wybiórczy. Nie jest to wielka wada, metod tych jest sporo, samych struktur grafowych, wykorzystywanych w asemblacji jest kilkadziesiąt (są wyliczone m.in. w pracy Kimia Behizadi at al., Graph theoretical Strategies in De Novo Assembling, 2022, doi: 10.1109/ACCESS.2022.3144113). Jednak wydaje mi się, że wypada obecnie wspomnieć o asemblacji hybrydowej, gdzie stosuje się jednocześnie krótkie i długie odczyty (np. Ilumina i Oxford Nanopore).

4 Podsumowanie

Stwierdzam, że **recenzowana rozprawa doktorska mgr Julii Herman-Iżyckiej spełnia warunki** określone w aktualnych przepisach i wnioskuję do Rady Naukowej Dyscyplin Matematyka i Informatyka Uniwersytetu Warszawskiego o dopuszczenie mgr Julii Herman-Iżyckiej do dalszych etapów przewodu doktorskiego.

Temat badawczy uważam za bardzo ciekawy i istotny. Osiągnięcia naukowe Kandydatki są wysokie, mgr Julia Herman-Iżycka posiada w swoim dorobku kilka prac współautorskich opublikowanych w renomowanych czasopismach naukowych i brała udział w grantach NCN i NCBiR.

Robert Nowel